

## S1 Appendix

Diana Catarino, Sergio Stefanni, Per Erik Jorde, Gui M. Menezes, Joan B. Company, Francis Neat and Halvor Knutsen 2017. The role of the Strait of Gibraltar in shaping the genetic structure of the Mediterranean grenadier, *Coryphaenoides mediterraneus*, between the Atlantic and Mediterranean Sea. PlosOne.

**Table A.** Preliminary analyses of pairwise  $F_{ST}$  values between sampling sites within Western (W1, W2 and W3) and Eastern (E1 and E2) Mediterranean. Based on the lack of genetic heterogeneity, all Western localities were grouped in one Western Mediterranean site (MED1) and the Eastern localities in one Eastern Mediterranean site (MED2) for all the analyses in the main manuscript. For all,  $P > 0.05$  for the exact  $G$  test of allele frequency heterogeneity over loci. Same order as in Table 1.

	W1MED	W2MED	W3MED	E1MED	E2MED
<b>W1MED</b>	—				
<b>W2MED</b>	-0.0090	—			
<b>W3MED</b>	-0.0049	-0.0155	—		
<b>E1MED</b>	0.0018	-0.0050	-0.0096	—	
<b>E2MED</b>	-0.0091	-0.0098	0.0130	0.0023	—

**Table B.** Nuclear microsatellite loci combined in multiplexes (Set) and annealing temperature ( $T_a$ ). Black lines separating the loci refer to how the loci were run (as multiplex or solo) during the PCR and the “Set” refers to the loci arrangement for the ABI 3130xl run. Estimated quantities:  $H_T$ , total heterozygosity;  $A$ , number of alleles scored per locus;  $R_s$ , mean allelic richness (minimum sample size of 95 individuals);  $F_{IS}$  deviations from Hardy-Weinberg equilibrium;  $F_{ST}$ , level of genetic differentiation among all sampled locations; overall<sup>1</sup> for nine loci; overall<sup>2</sup> for 7 loci (removing Crup7 and CaraA10). Significant at alpha = \*0.05, \*\* 0.01, \*\*\* 0.001, after FDR approach.

Loci	Set	$T_a$ (°C)	$H_T$	$A$	$R_s$	$F_{IS}$	$F_{ST}$
<b>Crup7</b>	<b>1</b>	56	0.926	35	28.1	<b>0.1094***</b>	<b>0.0269***</b>
<b>CaraA106a</b>	<b>1</b>	56	0.475	18	13.2	-0.0232	<b>0.0390***</b>
<b>Mbe03</b>			0.245	6	4.4	-0.0116	<b>0.0110*</b>
<b>CaraA109</b>	<b>2</b>	50	0.671	13	9.8	-0.0076	0.0001
<b>CaraC7</b>	<b>3</b>	58	0.468	12	8.8	-0.0407	<b>0.0134***</b>
<b>CaraA10</b>	<b>4</b>	56	0.164	5	2.8	<b>0.2356**</b>	0.0036
<b>CaraA102</b>			0.021	2	1.9	-0.0105	0.0036
<b>CaraB1</b>			0.044	4	3.0	0.1150	<b>0.0117**</b>
<b>CaraC1</b>			0.789	12	9.6	-0.0059	<b>0.0028*</b>
<b>Overall<sup>1</sup></b>			0.422			<b>0.0263***</b>	<b>0.0149***</b>
<b>Overall<sup>2</sup></b>			0.387			-0.0143	<b>0.0116***</b>

**Table C.** DIYABC summary statistics. Fields with an observed value correspond to the statistics used to build the reference table, the model choice analyses and to estimate the parameters. The observed value corresponds to the estimation in the real dataset. For each scenario the x indicates a summary statistic where the simulated value is similar to the observed value; the x followed by a \*, \*\* or \*\*\* indicates summary statistics where observed and simulated data were significantly different at 0.05, 0.01 or 0.001, respectively. Summary statistics not used in the reference table construction (i.e. statistics with no observed value calculated) were then used for model checking analysis.

Summary statistics		Observed value		Scenario 1		Scenario 2		Scenario 3		Scenario 4	
		ATL	MED	ATL	MED	ATL	MED	ATL	MED	ATL	MED
<b>One Sample</b>											
Microsats	Mean number of alleles	8.8889	7.1111	x**	x*	x**	x*	x	x*	x	x
	Mean genetic diversity	0.4423	0.3809	x**	x**	x**	x*	x*	x*	x*	x*
	Mean size variance										
	Mean Garza-Williamson's M	0.5063	0.6095	x***	x*	x***	x	x***	x***	x***	x**
	Number of haplotypes	4.0000	4.0000	x*	x	x*	x	x	x	x	x
mtDNA COI	Number of segregation sites	3.0000	5.0000	x*	x	x*	x	x	x	x	x
	Mean of pairwise differences										
	Variance of pairwise differences										
	Tajima's D	0.1190	-1.1593	x	x	x	x	x*	x	x*	x
	Private segregating sites										
mtDNA COI	Mean of numbers of the rarest nucleotide at segregating sites	7.3333	4.0000	x	x	x	x	x	x	x	x
	Variance of numbers of the rarest nucleotide at segregating sites										
<b>Two Sample</b>											
Microsats	Mean number of alleles	9.4444		x**		x**		x*		x*	
	Mean genetic diversity	0.4170		x***		x***		x**		x*	
	Mean size variance										
	Fst	0.0343		x		x		x		x	
	Classification index	0.8736	0.8115	x***	x***	x***	x***	x*	x**	x**	x**
mtDNA COI	Shared allele distance										
	(dμ) <sup>2</sup> distance										
	Number of haplotypes	8.0000		x		x		x		x	
	Number of segregation sites										
	Mean of pairwise differences (W)	0.5866		x*		x		x		x	
mtDNA COI	Mean of pairwise differences (B)										
	Fst (Hudson <i>et al.</i> , 1992)	0.6670		x*		x		x		x	

**Table D. Average genetic  $p$ -distances based on mtDNA COI for *C. mediterraneus* populations and other *Coryphaenoides* species.** Average  $p$ -distances are shown below the diagonal and standard error estimates are shown above the diagonal. The analysis involved 19 nucleotide sequences, each with 598 bp. Abbreviations: Cstr, *C. striatulus*; Cmur, *C. murrayi*; Ccar, *C. carapinus*; Cbre, *C. brevibarbis*; Cmed, *C. mediterraneus*; ATL, Atlantic; MED, Mediterranean.

	<b>Cstr</b>	<b>Cmur</b>	<b>Ccar</b>	<b>Cbre</b>	<b>Cmed ATL</b>	<b>Cmed MED</b>
<b>Cstr</b>	—	0.013	0.013	0.013	0.010	0.010
<b>Cmur</b>	0.115	—	0.013	0.013	0.012	0.012
<b>Ccar</b>	0.109	0.112	—	0.012	0.013	0.013
<b>Cbre</b>	0.110	0.120	0.105	—	0.013	0.013
<b>Cmed ATL</b>	0.057	0.092	0.106	0.112	—	0.002
<b>Cmed MED</b>	0.060	0.095	0.108	0.113	0.005	—

**Table E.** Hierarchical AMOVA analyses for the mtDNA COI fragment and the 9 microsatellite loci of the Mediterranean grenadier. Seven grouping criteria were tested and the highest support is found for those displaying the largest  $F_{CT}$  in relation to  $F_{SC}$ . In this perspective the best grouping option is shown in green for mtDNA and in orange for microsatellite markers. Sampled sites abbreviations are given in Table 1.

Grouping	Source of variation	mtDNA COI			Microsatellites		
		% of variance	$\Phi$ statistics	$P$	% of variance	$\Phi$ statistics	$P$
<b>Global panmixia</b> - total data set	Among populations	62.8	$\Phi_{ST} = 0.6279$	<b>&lt;0.001</b>	1.3	$F_{ST} = 0.0134$	<b>&lt;0.001</b>
	Within populations	37.2			98.7		
<b>Two regions - Atlantic vs. Mediterranean</b> (MAR1, MAR2, MAR3, ROC) (MED1, MED2)	Among groups	70	$\Phi_{CT} = 0.6999$	<b>&lt;0.05</b>	2.4	$F_{CT} = 0.0239$	<b>&lt;0.001</b>
	Among populations within groups	4.2	$\Phi_{SC} = 0.1392$	<b>&lt;0.001</b>	0.1	$F_{SC} = 0.0013$	0.221
	Within populations	25.8	$\Phi_{ST} = 0.7417$	<b>&lt;0.001</b>	97.5	$F_{ST} = 0.0252$	<b>&lt;0.001</b>
<b>Three regions</b> (MAR1, MAR2, MAR3) (ROC) (MED1, MED2)	Among groups	61.1	$\Phi_{CT} = 0.6108$	0.0506	1.4	$F_{CT} = 0.0141$	<b>0.003</b>
	Among populations within groups	5.8	$\Phi_{SC} = 0.1491$	<b>&lt;0.001</b>	0.1	$F_{SC} = 0.0014$	0.312
	Within populations	33.1	$\Phi_{ST} = 0.6689$	<b>&lt;0.001</b>	98.5	$F_{ST} = 0.0154$	<b>&lt;0.001</b>
(MAR1, ROC) (MAR2, MAR3) (MED1, MED2)	Among groups	66.3	$\Phi_{CT} = 0.6628$	<b>&lt;0.01</b>	1.6	$F_{CT} = 0.0160$	<b>&lt;0.001</b>
	Among populations within groups	1	$\Phi_{SC} = 0.0299$	0.1460	0.05	$F_{SC} = 0.0005$	0.4337
	Within populations	32.7	$\Phi_{ST} = 0.6728$	<b>&lt;0.001</b>	98.4	$F_{ST} = 0.0165$	<b>&lt;0.001</b>
(MAR1, ROC, MAR2) (MAR3) (MED1, MED2)	Among groups	65	$\Phi_{CT} = 0.6500$	<b>&lt;0.05</b>	1.7	$F_{CT} = 0.0172$	<b>&lt;0.001</b>
	Among populations within groups	2.9	$\Phi_{SC} = 0.6786$	<b>&lt;0.01</b>	0	$F_{SC} = 0.0000$	0.4915
	Within populations	32.1	$\Phi_{ST} = 0.0818$	<b>&lt;0.001</b>	98.3	$F_{ST} = 0.0172$	<b>&lt;0.001</b>

**Table F.** Pairwise  $F_{ST}$  calculated using Arlequin.

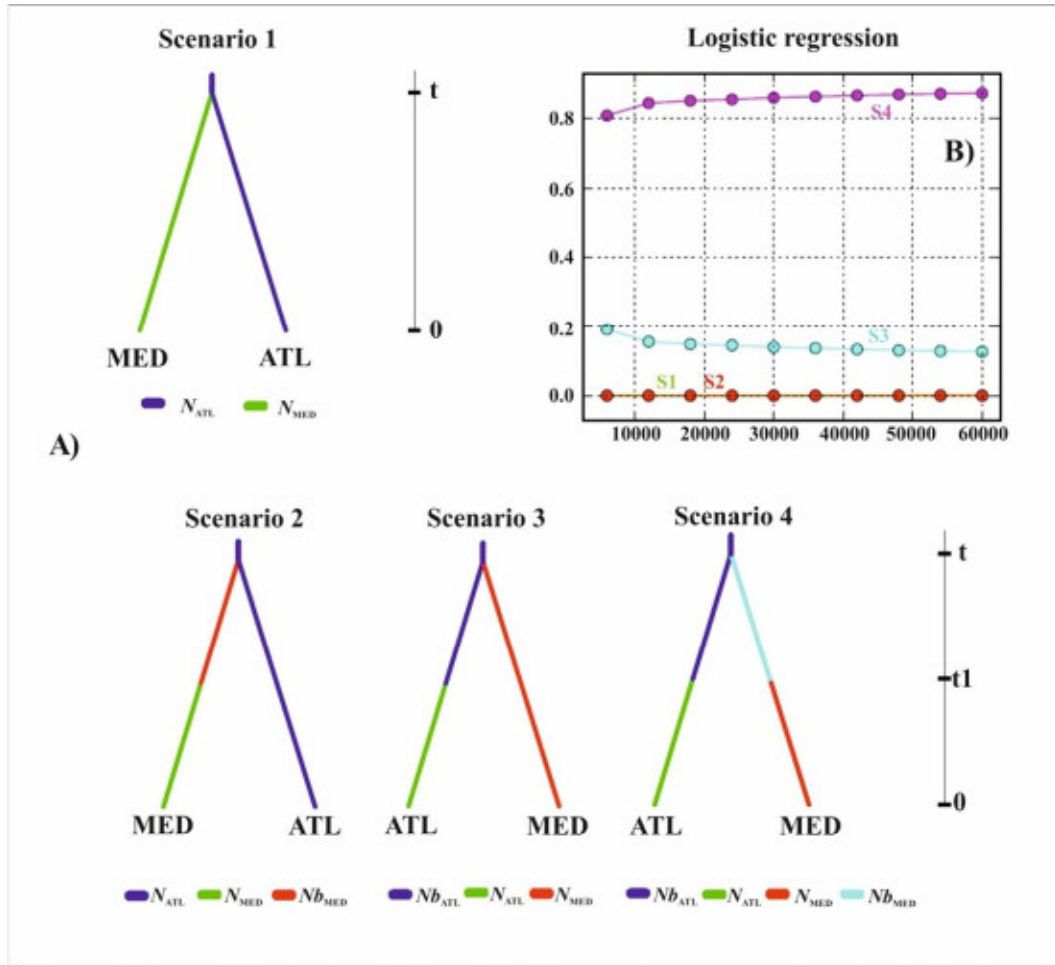
	MAR1	MAR2	MAR3	ROC	MED1	MED2
MAR1	—					
MAR2	-0.0012	—				
MAR3	0.0026	-0.0023	—			
ROC	0.0010	-0.0024	0.0017	—		
MED1	<b>0.0199***</b>	<b>0.0426***</b>	<b>0.0312***</b>	<b>0.0218***</b>	—	
MED2	<b>0.0260***</b>	<b>0.0449**</b>	<b>0.0371***</b>	<b>0.0264***</b>	0.0011	—

**Table G.** Neutrality test for *C. mediterraneus* mtDNA COI sequences. \*  $P < 0.01$ .

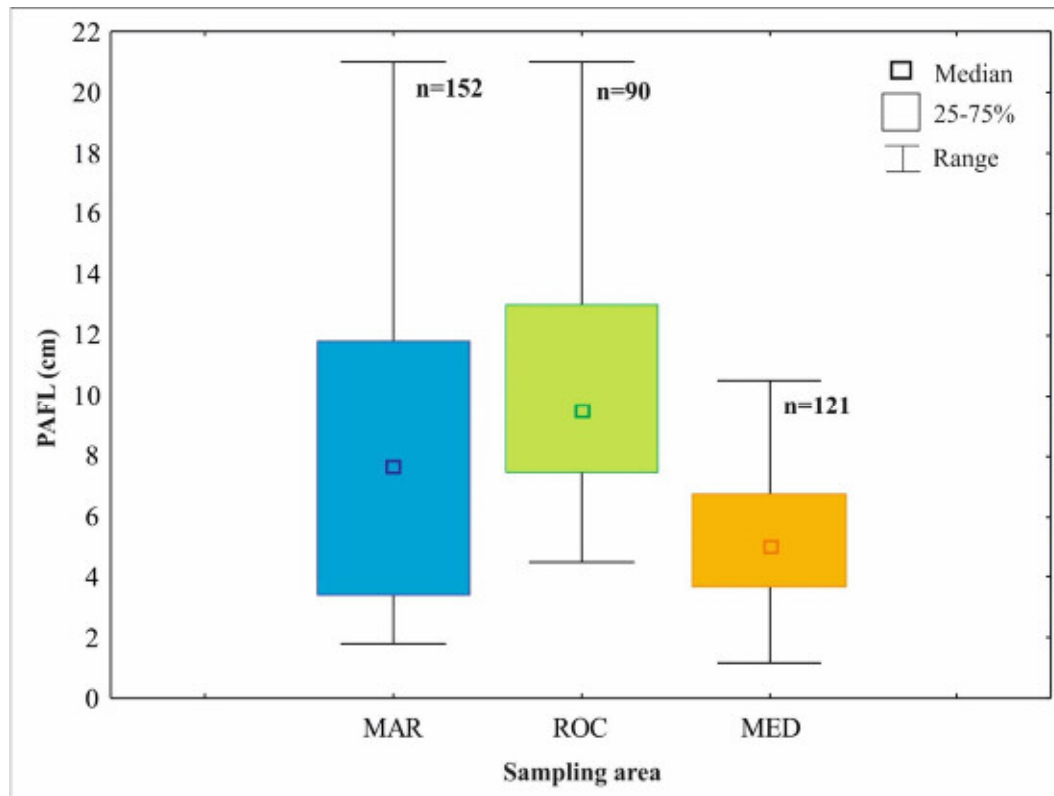
	MAR1/ROC	MAR2/MAR3	Mediterranean
Number of individuals	79	48	66
Tajima's $D$	-1.311	0.119	-1.159
Fu's $F_S$	<b>-2.945*</b>	-0.019	-0.559
Rozas's $R_2$	0.057	0.119	0.067

**Table H.** Bottleneck tests results for *C. mediterraneus* samples. P refers to the probability of each test. Different combinations of percentage of stepwise mutations and variance were tested in TPM model: **1)** 70, 10; **2)** 70, 30; **3)** 90, 10; **4)** 90, 30, for the percentage and variance, respectively.

Population	P (H deficient)				
	SMM	TPM			
		1	2	3	4
<b>Atlantic</b>	0.00098	0.00098	0.00098	0.00098	0.00098
<b>Mediterranean</b>	0.00098	0.00488	0.00684	0.00195	0.00195

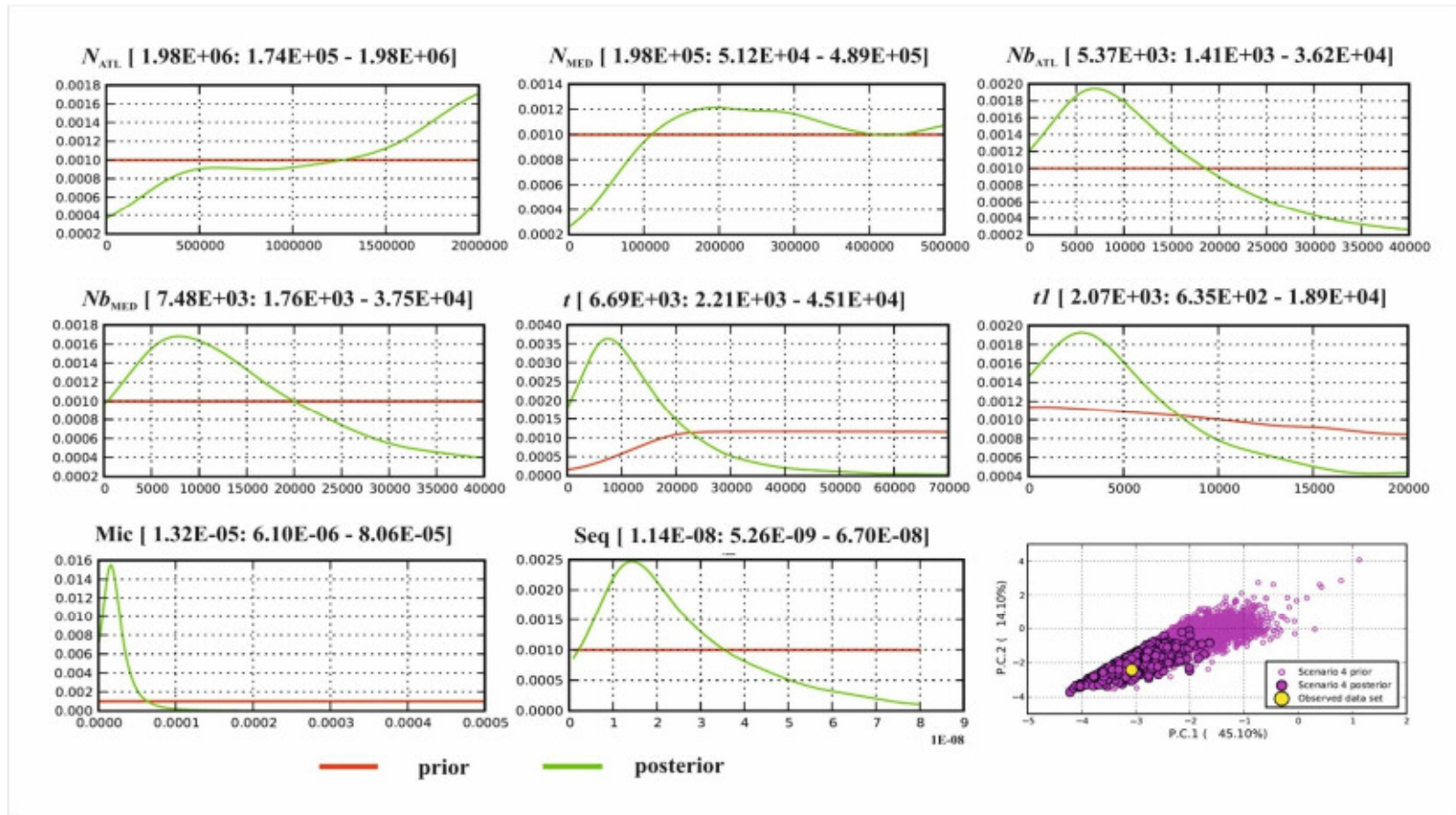


**Figure A.** DIYABC scenarios under consideration for *C. mediterraneus* populations. **A)** schematic representation of the four alternative scenarios used for Approximate Bayesian computation simulations. All scenarios assume that at present time  $t(0)$ , there are two populations Atlantic (ATL) and Mediterranean (MED) and that they diverged from a single population in the past at  $t$  generations.  $N_{ATL}$  and  $N_{MED}$  indicate periods where population size was allowed to change. For detailed information on each scenario see Material and Methods section; **B)** Relative posterior probability of the demographic scenarios estimated using the logistic regression approach. Scenario 1 and 2 are overlapping at the bottom of the graph (see Table 4 for probability values and 95% CI).



**Figure B.** Resume of the length data based on the indicated number (*n*) of specimens collected in each of the main studied areas. Length was recorded as the pre-anal fin length (PAFL). Sampling area abbreviations are given in Table 1.





**Figure C.** DIYABC posterior probability distributions (green lines) for each parameter estimated under scenario 4. Besides the distributions it is also shown the modal values (95% Credible Interval) for Atlantic and Mediterranean  $N_e$  ( $N_{ATL}$  and  $N_{MED}$ ), Atlantic and Mediterranean  $N_e$  after divergence ( $Nb_{ATL}$  and  $Nb_{MED}$ ), time since divergence ( $t$ ),  $tI$  and the distributions for the mutation rates for microsatellites ( $Mic$ ) and mtDNA ( $Seq$ ). The last image refer to the “model checking” analyses using a PCA to evaluate the model fitting of the dataset, using a different set of summary statistics from those used to estimate the parameters.